A genome-wide association study for diabetic nephropathy genes in African Americans

Caitrin W. McDonough[1,2,3], Nicholette D. Palmer[2,3,4], Pamela J. Hicks[2,3,4], Bong H. Roh[2,3,4], S. Sandy An[2,3,4], Jessica N. Cooke[1,2,3], Jessica M. Hester[2,3,5], Maria R. Wing[2,3,5], Meredith A. Bostrom[2,3,4], Megan E. Rudock[2,3], Joshua P. Lewis[2,3,5], Matthew E. Talbert[1,2,3], Rebecca A. Blevins[2,3,4], Lingyi Lu[6], Maggie C.Y. Ng[2,3,7], Michele M. Sale[8], Jasmin Divers[6], Carl D. Langefeld[6], Barry I. Freedman[9] and Donald W. Bowden[2,3,4,9]

[1]Program in Molecular Medicine and Translational Science, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [2]Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [3]Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [4]Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [5]Program in Molecular Genetics and Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [6]Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [7]Department of Pediatrics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA; [8]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, Virginia, USA and [9]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA

A genome-wide association study was performed using the Affymetrix 6.0 chip to identify genes associated with diabetic nephropathy in African Americans. Association analysis was performed adjusting for admixture in 965 type 2 diabetic African American patients with end-stage renal disease (ESRD) and in 1029 African Americans without type 2 diabetes or kidney disease as controls. The top 724 single nucleotide polymorphisms (SNPs) with evidence of association to diabetic nephropathy were then genotyped in a replication sample of an additional 709 type 2 diabetes-ESRD patients and 690 controls. SNPs with evidence of association in both the original and replication studies were tested in additional African American cohorts consisting of 1246 patients with type 2 diabetes without kidney disease and 1216 with non-diabetic ESRD to differentiate candidate loci for type 2 diabetes-ESRD, type 2 diabetes, and/or all-cause ESRD. Twenty-five SNPs were significantly associated with type 2 diabetes-ESRD in the genome-wide association and initial replication. Although genome-wide significance with type 2 diabetes was not found for any of these 25 SNPs, several genes, including *RPS12*, *LIMK2*, and *SFI1* are strong candidates for diabetic nephropathy. A combined analysis of all 2890 patients with ESRD showed significant association SNPs in *LIMK2* and *SFI1* suggesting that they also contribute to all-cause ESRD. Thus, our results suggest that multiple loci underlie susceptibility to kidney disease in African Americans with type 2 diabetes and some may also contribute to all-cause ESRD.

Kidney International (2011) **79,** 563–572; doi:10.1038/ki.2010.467; published online 8 December 2010

KEYWORDS: diabetic nephropathy; end-stage renal disease; ethnic minority; human genetics; polymorphisms

Correspondence: *Donald W. Bowden, Department of Biochemistry, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, North Carolina 27157, USA. E-mail: dbowden@wfubmc.edu*

Received 14 May 2010; revised 23 September 2010; accepted 28 September 2010; published online 8 December 2010

Diabetes-associated nephropathy is the most common cause of end-stage renal disease (ESRD) in the United States, accounting for 44.6% of incident cases.[1] African Americans (AAs) have a disproportionately high risk for developing diabetic nephropathy (DN). Compared with Caucasian Americans (CAs), AAs have a 3.7-fold greater incidence rate of developing ESRD and at least a 1.9-fold greater incidence rate than do other racial and ethnic minorities in the United States.[1] Many studies have shown that there is a genetic component to ESRD as reviewed by Bowden.[2] Familial aggregation of DN and diabetic ESRD has been demonstrated in CAs[3–5] and AAs.[6] Clustering occurred in these families without significant differences in glycemic control.[4] However, marked racial and ethnic disparities in familial clustering exist. CAs who have a close relative with ESRD face a 2.7-fold increased risk of developing ESRD,[5] whereas AAs who have a close relative with ESRD have a 9-fold increased risk of developing ESRD.[6] This significant difference in rates of renal complications between CAs and AAs is observed after controlling for differences in socioeconomic status.[5,6]

Several studies have attempted to detect genetic variants influencing the risk of DN and diabetic ESRD. The first genome-wide association study (GWAS) for DN was a low-density (80K single-nucleotide polymorphisms (SNPs)) gene-based study performed in a Japanese population.[7] This

was followed by a study using a DNA pooling method investigating 115K SNPs in Pima Indians with DN.[8] More recently, a GWAS was conducted for type 1 diabetes-associated nephropathy in a CA population,[9] and multiple studies have assessed for association with chronic kidney disease and glomerular filtration rate in nondiabetic populations of European ancestry.[10–12] There have been no previous reports of GWASs in AAs with type 2 diabetes mellitus (T2DM)-ESRD. In this study, we report the first GWAS investigating 832K SNPs for association with T2DM-ESRD in AAs.

RESULTS

Clinical characteristics of study samples

The clinical characteristics of study samples used in the GWAS, replication, and trait discrimination phases are shown in Table 1. The GWAS and replication populations are broadly similar. In both groups, the age at enrollment for T2DM-ESRD subjects is older than that for the control groups. However, the age at enrollment for the control groups in the GWAS and replication phases is older than the age of T2DM diagnosis in T2DM-ESRD and T2DM subjects. All of the case groups with T2DM (T2DM-ESRD and T2DM) have a higher proportion of females, possibly reflecting the increased prevalence of T2DM among AA women,[13] participation bias, and survival. On average, all of the groups were overweight or obese at the time of enrollment. Subjects with ESRD lacking T2DM (non-T2DM-ESRD subjects) had the lowest average body mass index (27.0 kg/m^2, Table 1), and T2DM subjects without nephropathy (T2DM) had the highest average body mass index (33.5 kg/m^2, Table 1).

Genome-wide association study

After the application of SNP and sample quality control metrics, 832,357 autosomal SNPs were analyzed in 965 AA T2DM-ESRD case subjects and in 1029 AA nondiabetic, nonnephropathy controls. A summary of the association results is shown in Figure 1 and the corresponding quantile–quantile plot is shown in Supplementary Figure S1 online. The results shown are adjusted for admixture; however, the primary inferences remain the same adjusting for admixture, age, and gender. The top hit was rs5750250 located on chromosome 22 in the *MYH9* (nonmuscle myosin heavy chain 9) gene ($P = 3.00 \times 10^{-7}$, Figure 1). This gene has been previously associated with both nondiabetic and diabetic forms of ESRD.[14–17] In total, there were 126 SNPs with *P*-values $< 1.0 \times 10^{-4}$ (Figure 1). The flow of the study through the GWAS, replication, combined and trait discrimination phases is outlined in Table 2.

Replication and combined analysis of T2DM-ESRD cases and nondiabetic, nonnephropathy controls

In an effort to replicate the GWAS results, 724 top-scoring SNPs were genotyped in an independent sample of 709 AA T2DM-ESRD cases and 690 AA nondiabetic, nonnephropathy controls (study design, Table 2; Results, Supplementary Table S1 online). The 724 SNPs that were selected for testing in the replication sample were SNPs with the strongest *P*-values for association and with high quality scores for genotyping, that is, missing rate < 0.02 (or missing rate < 0.05, but the missing rate between cases and controls was not significantly different). In addition, *P*-values for Hardy–Weinberg proportions were > 0.0001 for cases and > 0.01 for controls. SNPs were prioritized on significance of the additive genetic model unless there was *a priori* evidence for follow-up, for example, *MYH9*. In this replication analysis, 67 SNPs showed nominal evidence of replication: additive *P*-value < 0.05 with association in the same direction (Supplementary Table S1 online). Table 3 summarizes the association results for 25 SNPs at 19 potential T2DM-ESRD loci. T2DM-ESRD loci were identified during trait discrimination analyses as shown in Table 2 and were based on the following criteria: (1) associated in the replication phase, (2) associated in the T2DM-ESRD versus T2DM, nonnephropathy comparison (Table 4), and (3) showed no association or nominal association ($P > 0.01$) in the T2DM, nonnephropathy versus controls comparison (Table 4). Table 3 shows *P*-values for association in the GWAS, the replication sample, and the combined cohort: 1674 T2DM-ESRD cases and 1719 nondiabetic, nonnephropathy controls. No SNP reached genome-wide significance ($P \leqslant 5 \times 10^{-8}$), *P*-values ranged from 1.24×10^{-4} to 7.04×10^{-7} (Table 3, combined analysis).

The strongest association in the combined analysis (GWAS + replication) was with rs6930576 ($P = 7.04 \times 10^{-7}$, odds ratio (OR) (95% confidence interval (95% CI)) = 1.31 (1.18–1.45); Table 3, Supplementary Figure S2a online). SNP

Table 1 | Clinical characteristics of study samples

	GWAS		Replication		Trait discrimination	
	T2DM-ESRD	Controls	T2DM-ESRD	Controls	T2DM	Non-T2DM-ESRD
n	965	1029	709	690	1246	1216
Female (%)	61.20	57.30	55.70	51.30	64.00	44.70
Age at enrollment (years)	61.6 ± 10.5	49.0 ± 11.9	60.2 ± 10.4	48.5 ± 12.8	57.2 ± 11.7	53.0 ± 14.5
Age at T2D diagnosis (years)	41.6 ± 12.4	—	39.4 ± 12.5	—	46.1 ± 12.6	—
Age at ESRD diagnosis (years)	58.0 ± 10.9	—	56.7 ± 10.9	—	—	47.7 ± 15.5
T2D to ESRD duration (years)	16.2 ± 10.9	—	20.4 ± 10.5	—	—	—
BMI (kg/m^2)	29.7 ± 7.0	30.0 ± 7.0	29.8 ± 6.9	29.4 ± 7.6	33.5 ± 7.6	27.0 ± 7.0

Abbreviations: BMI, body mass index; ESRD, end-stage renal disease; GWAS, genome-wide association study; T2DM, type 2 diabetes mellitus.
Values are presented as trait mean and s.d.

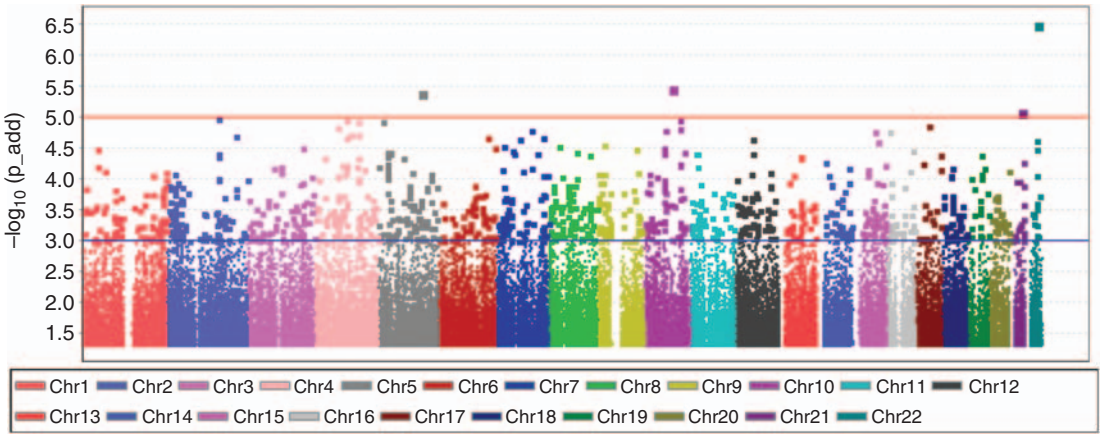

Figure 1 | Genome-wide association study results. The top hit was rs5750250 located on chromosome 22 in the *MYH9* gene ($P = 3.00 \times 10^{-7}$). Results are adjusted for admixture using PC1 as a covariate in the analysis. *P*-values are shown under the additive model. Chr, chromosome; MYH9, nonmuscle myosin heavy chain 9; PC1, first principal component.

Table 2 | Study design and progression

Step	Stage	Sample			Description
		Total	Cases	Controls	
1	GWAS	1994	965 T2DM-ESRD cases	1029 nondiabetic, nonnephropathy controls	GWAS scan with 832,357 SNPs
2	Replication	1399	709 T2DM-ESRD cases	690 nondiabetic, nonnephropathy controls	Test 724 SNPs from step 1 for replication
3	Combined analysis (GWAS+replication)	3393	1674 T2DM-ESRD cases	1719 nondiabetic, nonnephropathy controls	Identify SNPs with the strongest evidence of association with T2DM-ESRD
4	Trait discrimination test for T2DM or T2DM-ESRD	2920	1674 T2DM-ESRD cases	1246 T2DM (lacking nephropathy) cases	Test 67 SNPs from step 3 to differentiate between T2DM-ESRD and T2DM loci
5	Test for non-T2DM-ESRD	2935	1216 non-T2DM-ESRD cases	1719 nondiabetic, nonnephropathy controls	Test 67 SNPs from step 3 for association with all-cause ESRD
6	Test for all-cause ESRD	4609	2890 ESRD cases (1674 T2DM-ESRD and 1216 non-T2DM-ESRD)	1719 nondiabetic, nonnephropathy controls	Test 67 SNPs from step 3 for association with all-cause ESRD in the largest sample set

Abbreviations: ESRD, end-stage renal disease; GWAS, genome-wide association study; SNP, single-nucleotide polymorphism; T2DM, type 2 diabetes mellitus.

rs6930576 is located in intron 1 of the SAM and SH3 domain containing 1 (*SASH1*) gene. Six other hits were detected in four regions with combined *P*-values $< 1.0 \times 10^{-5}$. SNP rs7769051 is located 8 kb upstream of the ribosomal S12 (*RPS12*) gene ($P = 2.2 \times 10^{-6}$, OR (95% CI) = 1.28 (1.16–1.42); Table 3, Supplementary Figure S2b online). SNP rs773506 ($P = 6.45 \times 10^{-6}$, OR (95% CI) = 0.76 (0.67–0.85); Table 3, Supplementary Figure S2c online) is located 600 bp 3′ of *AUH* (AU RNA-binding protein/enoyl-Coenzyme A). Next, SNP rs2358944 ($P = 3.54 \times 10^{-6}$, OR (95% CI) = 0.75 (0.67–0.85); Table 3, Supplementary Figure S2d online) is located 38 kb from the noncoding RNA, ribosomal protein SA pseudogene 52 (*RPSAP52*). This SNP, rs2358944, is also located near two coding genes: *HMGA2*, high mobility group AT-hook 2 (101 kb upstream of the variant), and *MSRB3*, methionine sulfoxide reductase B3 (257 kb downstream of the marker). Finally, three hits were located in the *LIMK2* (LIM domain kinase 2) —*SFI1* (Sfi1 homolog, spindle assembly-associated (yeast)) region: rs2106294 and rs4820043 are both located in intron 2 of *LIMK2* ($P = 4.11 \times 10^{-6}$ and $P = 5.07 \times 10^{-6}$, OR (95% CI) = 0.57 (0.45–0.72) and OR (95% CI) = 0.57 (0.45–0.73), respectively; Table 3, Supplementary Figure S2e online), and rs5749286 is located in intron 1 of *SFI1* ($P = 9.79 \times 10^{-6}$, OR (95% CI) = 0.60 (0.47–0.75); Table 3, Supplementary Figure S2e online). The *MYH9* SNP seen in the initial GWAS, rs5750250, did not replicate ($P = 0.0594$, Supplementary Table S2 online) and had a combined *P*-value of 1.66×10^{-7} (Supplementary Table S2 online).

Discrimination between T2DM-ESRD, T2DM, and all-cause ESRD

The 67 SNPs that showed evidence of replication between T2DM-ESRD cases and nonnephropathy, nondiabetic controls were genotyped in an additional 1246 AA T2DM, nonnephropathy cases and 1216 non-T2DM-ESRD cases to differentiate between T2DM-ESRD, T2DM, and all-cause ESRD loci (Table 2; results shown in Supplementary Table S3

Table 3 | Diabetic nephropathy candidate loci

Locus					GWAS T2DM-ESRD (*n*=965) controls (*n*=1029)			Replication T2DM-ESRD (*n*=709) controls (*n*=690)			Combined T2DM-ESRD (*n*=1674) controls (*n*=1719)	
Marker	Chr.	Position	Nearest gene(s)	Minor allele (major allele)	MAF T2DM-ESRD	MAF controls	Additive *P*-value	MAF T2DM-ESRD	MAF controls	Additive *P*-value	Additive *P*-value	OR (95% CI)
rs10888287	1	246,184,039	*OR2L13*	G(T)	0.439	0.503	5.03E-04	0.444	0.495	0.0482	5.97E-05	0.82 (0.75–0.90)
rs4260465	3	164,231,046	No gene	G(C)	0.364	0.414	4.31E-04	0.381	0.421	0.0382	4.50E-05	0.81 (0.74–0.90)
rs11730446	4	96,382,267	*UNC5C*	G(A)	0.459	0.403	6.23E-04	0.459	0.424	0.0465	1.24E-04	1.21 (1.10–1.33)
rs891382	4	147,371,790	*SLC10A7/LSM6*	G(A)	0.085	0.059	3.19E-04	0.081	0.065	0.0183	2.42E-05	1.49 (1.24–1.80)
rs7697691	4	182,853,951	No gene	G(A)	0.243	0.285	0.0016	0.243	0.285	0.0127	7.37E-05	0.80 (0.72–0.89)
rs3822908	6	97,706,333	*C6orf167*	C(T)	0.125	0.162	3.63E-04	0.121	0.147	0.0462	4.52E-05	0.75 (0.65–0.86)
rs7760831	6	102,071,536	*GRIK2*	C(T)	0.176	0.222	2.55E-04	0.193	0.224	0.0455	3.48E-05	0.77 (0.69–0.87)
rs208865	6	130,114,249	*C6orf191/ARHGAP18*	T(C)	0.118	0.082	2.14E-04	0.115	0.091	0.0364	2.90E-05	1.41 (1.20–1.66)
rs9493454	6	133,186,322	*RPS12*	C(A)	0.439	0.376	4.91E-04	0.453	0.395	0.0089	1.72E-05	1.24 (1.13–1.37)
rs7769051	6	133,188,489	*RPS12*	A(C)	0.349	0.289	4.06E-04	0.362	0.3	0.0012	2.20E-06	1.28 (1.16–1.42)
rs6930576	6	148,746,647	*SASH1*	A(G)	0.337	0.276	1.87E-05	0.319	0.274	0.0099	7.04E-07	1.31 (1.18–1.45)
rs773506	9	93,015,292	*AUH*	A(G)	0.179	0.233	3.27E-04	0.174	0.223	0.0130	6.45E-06	0.76 (0.67–0.85)
rs11175885	12	64,400,897	*C12orf66/TMEM5*	G(A)	0.181	0.234	0.0013	0.19	0.249	0.0106	2.18E-05	0.77 (0.68–0.87)
rs2358944	12	64,403,825	*MSRB3/HMGA2*	A(G)	0.178	0.232	9.48E-04	0.188	0.256	0.0018	3.54E-06	0.75 (0.67–0.85)
rs2904532	12	65,582,311	*GRIP1/CAND1*	G(A)	0.141	0.183	8.17E-04	0.148	0.184	0.0172	4.87E-05	0.76 (0.67–0.87)
rs11176482	12	65,582,814	*GRIP1/CAND1*	C(T)	0.141	0.184	7.08E-04	0.147	0.182	0.0202	4.79E-05	0.76 (0.67–0.87)
rs12302041	12	76,654,356	*NAV3*	A(G)	0.085	0.055	3.09E-04	0.088	0.06	0.0149	1.15E-05	1.54 (1.27–1.87)
rs1978243	14	86,596,160	No gene	T(C)	0.22	0.274	7.53E-04	0.221	0.274	0.0105	1.61E-05	0.78 (0.70–0.87)
rs6494387	15	61,132,579	*TPM1*	A(G)	0.367	0.311	2.32E-04	0.358	0.32	0.0209	1.80E-05	1.25 (1.13–1.39)
rs1034589	22	29,909,233	*RNF185*	C(T)	0.036	0.062	0.0010	0.033	0.065	0.0055	1.06E-05	0.59 (0.46–0.74)
rs2106294	22	29,975,759	*LIMK2*	C(T)	0.035	0.062	7.49E-04	0.031	0.064	0.0028	4.11E-06	0.57 (0.45–0.72)
rs4820043	22	29,977,094	*LIMK2*	A(G)	0.036	0.062	0.0010	0.031	0.065	0.0026	5.07E-06	0.57 (0.45–0.73)
rs5749286	22	30,230,359	*SFI1*	A(C)	0.038	0.065	0.0013	0.036	0.068	0.0030	9.79E-06	0.60 (0.47–0.75)
rs16996381	22	34,841,225	*APOL3*	G(A)	0.336	0.391	0.0018	0.371	0.432	0.0191	1.06E-04	0.82 (0.75–0.91)
rs735853	22	35,009,161	*MYH9*	G(C)	0.082	0.123	4.59E-04	0.079	0.118	0.0091	1.20E-05	0.69 (0.59–0.82)

Abbreviations: Chr., chromosome; ESRD, end-stage renal disease; GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; T2DM, type 2 diabetes mellitus; 95% CI, 95% confidence interval.
GWAS, replication, and combined (GWAS+replication) *P*-values for diabetic nephropathy candidate loci across the genome.

Table 4 | Diabetic nephropathy candidate loci: trait discrimination

Locus				T2DM-ESRD (*n*=1674) versus T2DM (*n*=1246)		T2DM (*n*=1246) versus controls (*n*=1719)	
Marker	Chr.	Position	Nearest gene(s)	Additive *P*-value	OR (95% CI)	Additive *P*-value	OR (95% CI)
rs10888287	1	246,184,039	*OR2L13*	0.0361	0.89 (0.80–0.99)	0.1487	0.93 (0.84–1.03)
rs4260465	3	164,231,046	No gene	2.63E-05	0.80 (0.72–0.89)	0.5177	1.03 (0.93–1.15)
rs11730446	4	96,382,267	*UNC5C*	0.0068	1.16 (1.04–1.29)	0.3319	1.05 (0.95–1.17)
rs891382	4	147,371,790	*SLC10A7/LSM6*	1.86E-04	1.49 (1.21–1.83)	0.8984	1.01 (0.82–1.25)
rs7697691	4	182,853,951	No gene	0.0046	0.84 (0.75–0.95)	0.3664	0.95 (0.84–1.07)
rs3822908	6	97,706,333	*C6orf167*	0.0336	0.84 (0.72–0.99)	0.1979	0.91 (0.78–1.05)
rs7760831	6	102,071,536	*GRIK2*	0.0494	0.87 (0.76–1.00)	0.0587	0.88 (0.78–1.00)
rs208865	6	130,114,249	*C6orf191/ARHGAP18*	0.0220	1.23 (1.03–1.46)	0.1259	1.15 (0.96–1.37)
rs9493454	6	133,186,322	*RPS12*	8.79E-04	1.20 (1.08–1.34)	0.5079	1.04 (0.93–1.16)
rs7769051	6	133,188,489	*RPS12*	0.0036	1.18 (1.06–1.32)	0.1614	1.09 (0.97–1.22)
rs6930576	6	148,746,647	*SASH1*	0.0053	1.18 (1.05–1.32)	0.0621	1.12 (0.99–1.25)
rs773506	9	93,015,292	*AUH*	2.57E-04	0.78 (0.69–0.89)	0.7484	0.98 (0.87–1.11)
rs11175885	12	64,400,897	*C12orf66/TMEM5*	0.0226	0.86 (0.75–0.98)	0.1068	0.90 (0.79–1.02)
rs2358944	12	64,403,825	*MSRB3/HMGA2*	0.0047	0.83 (0.72–0.94)	0.1532	0.91 (0.80–1.03)
rs2904532	12	65,582,311	*GRIP1/CAND1*	0.0047	0.81 (0.71–0.94)	0.3862	0.94 (0.82–1.08)
rs11176482	12	65,582,814	*GRIP1/CAND1*	0.0070	0.82 (0.71–0.95)	0.3052	0.93 (0.81–1.07)
rs12302041	12	76,654,356	*NAV3*	0.0226	1.27 (1.03–1.55)	0.0924	1.21 (0.97–1.50)
rs1978243	14	86,596,160	No gene	0.0332	0.87 (0.77–0.99)	0.0743	0.90 (0.79–1.01)
rs6494387	15	61,132,579	*TPM1*	0.0155	1.15 (1.03–1.29)	0.1683	1.08 (0.97–1.21)
rs1034589	22	29,909,233	*RNF185*	0.0274	0.74 (0.57–0.97)	0.0438	0.79 (0.63–0.99)
rs2106294	22	29,975,759	*LIMK2*	0.0167	0.72 (0.55–0.94)	0.0411	0.79 (0.63–0.99)
rs4820043	22	29,977,094	*LIMK2*	0.0162	0.72 (0.55–0.94)	0.0460	0.79 (0.63–1.00)
rs5749286	22	30,230,359	*SFI1*	0.0274	0.75 (0.58–0.97)	0.0344	0.78 (0.62–0.98)
rs16996381	22	34,841,225	*APOL3*	1.76E-04	0.82 (0.74–0.91)	0.8152	1.01 (0.91–1.12)
rs735853	22	35,009,161	*MYH9*	8.57E-04	0.73 (0.61–0.88)	0.4515	0.94 (0.80–1.11)

Abbreviations: Chr., chromosome; ESRD, end-stage renal disease; OR, odds ratio; T2DM, type 2 diabetes mellitus; 95% CI, 95% confidence interval.
Comparison of T2DM-ESRD versus T2DM, nonnephropathy subjects and T2DM, nonnephropathy subjects versus nondiabetic normal controls in diabetic nephropathy candidate loci across the genome.

online). Thus, T2DM-ESRD SNPs should have allele frequency differences when compared with cases with T2DM alone (lacking nephropathy). When the 25 T2DM-ESRD candidate SNPs were compared between T2DM-ESRD cases and T2DM subjects without nephropathy, there were significant differences between these two groups (*P*-values ranging from 0.049 to 2.63×10^{-5}, Table 4), suggesting that these are T2DM-ESRD loci rather than T2DM loci. SNPs in *RPS12* (rs9493454) and *AUH* (rs7735506) were still strongly associated in this test ($P = 8.79 \times 10^{-4}$ and $P = 2.57 \times 10^{-4}$, OR (95% CI) = 1.20 (1.08–1.34) and OR (95% CI) = 0.78 (0.69–0.89), respectively; Table 4).

As a second test, allele frequencies of these 25 SNPs in T2DM, nonnephropathy subjects were compared with those in normal, nondiabetic, nonnephropathy controls (Table 4; control data from the GWAS and replication control samples). True T2DM loci would be expected to show evidence of association in this analysis. Although the four SNPs in the *LIMK2-SFI1* region (namely rs1034589, rs2106294, rs4820043, and rs5749286) showed nominal evidence of association with *P*-values ranging from 0.034 to 0.046 (OR ranging from 0.78 to 0.79, Table 4), these *P*-values are orders of magnitude lower than comparing T2DM-ESRD with controls (Table 3; combined analysis). This further supports these SNPs as T2DM-ESRD loci and not as T2DM loci.

The 25 SNPs were also investigated for association with all-cause ESRD. In this analysis, cases consisted of 1674 T2DM-ESRD subjects (from the GWAS and replication phases) and 1216 nondiabetic ESRD subjects (predominantly ESRD attributed to hypertension and chronic glomerular diseases). These 2890 cases with ESRD were compared with the combined set of nondiabetic, nonnephropathy controls ($n = 1719$, from the GWAS and replication phases). Five variants in two genomic regions achieved genome-wide significance in this analysis ($P \leqslant 5.0 \times 10^{-8}$, Table 5). The *LIMK2-SFI1* region contained four of these SNPs. First, rs1034589, located in ring finger protein 185 (*RNF185*), was associated with all-cause ESRD ($P = 2.49 \times 10^{-8}$, OR (95% CI) = 0.56 (0.46–0.69); Table 5, Figure 2a). Two SNPs showed evidence of association in the *LIMK2* gene, namely rs2106294 and rs4820043 ($P = 3.28 \times 10^{-8}$ and $P = 3.04 \times 10^{-8}$, OR (95% CI) = 0.56 (0.46–0.69) and OR (95% CI) = 0.56 (0.46–0.69), respectively; Table 5, Figure 2a). The last SNP that was significantly associated in this region was rs5749286, located in *SFI1* ($P = 2.96 \times 10^{-11}$, OR (95% CI) = 0.51 (0.41–0.62); Table 5, Figure 2a). Finally, SNP rs735853, in *MYH9* was also significantly associated with all-cause ESRD ($P = 4.77 \times 10^{-10}$, OR (95% CI) = 0.63 (0.54–0.73); Table 5, Figure 2b). All markers were also examined for association with nondiabetic ESRD compared with controls (Table 5; control data from the GWAS and replication control samples). In all, 12 markers showed association with nondiabetic ESRD (*P*-values ranging from 0.0364 to 1.97×10^{-10}, ORs ranging from 0.38 to 0.86 and from 1.15 to 1.31, Table 5). However, all of these variants were more

Table 5 | Diabetic nephropathy candidate loci: all-cause ESRD

Locus				All-cause ESRD (*n*=2890) versus controls (*n*=1719)		non-T2DM-ESRD (*n*=1216) versus controls (*n*=1719)	
Marker	Chr.	Position	Nearest gene(s)	Additive *P*-value	OR (95% CI)	Additive *P*-value	OR (95% CI)
rs10888287	1	246,184,039	*OR2L13*	5.53E-05	0.84 (0.77–0.91)	0.0057	0.86 (0.78–0.96)
rs4260465	3	164,231,046	No gene	0.0027	0.88 (0.81–0.96)	0.4548	0.96 (0.87–1.07)
rs11730446	4	96,382,267	*UNC5C*	8.25E-05	1.19 (1.09–1.30)	0.0085	1.15 (1.04–1.28)
rs891382	4	147,371,790	*SLC10A7/LSM6*	2.56E-04	1.37 (1.16–1.63)	0.0859	1.20 (0.97–1.48)
rs7697691	4	182,853,951	No gene	0.0036	0.87 (0.79–0.95)	0.6059	0.97 (0.86–1.09)
rs3822908	6	97,706,333	*C6orf167*	2.43E-04	0.80 (0.71–0.90)	0.0364	0.85 (0.73–0.99)
rs7760831	6	102,071,536	*GRIK2*	1.62E-06	0.77 (0.69–0.86)	1.43E-04	0.77 (0.68–0.88)
rs208865	6	130,114,249	*C6orf191/ARHGAP18*	2.60E-05	1.37 (1.18–1.59)	0.0025	1.31 (1.10–1.56)
rs9493454	6	133,186,322	*RPS12*	2.35E-05	1.21 (1.11–1.32)	0.2134	1.07 (0.96–1.19)
rs7769051	6	133,188,489	*RPS12*	5.85E-06	1.24 (1.13–1.36)	0.1958	1.07 (0.96–1.19)
rs6930576	6	148,746,647	*SASH1*	4.76E-07	1.27 (1.16–1.40)	4.80E-04	1.23 (1.09–1.38)
rs773506	9	93,015,292	*AUH*	1.40E-04	0.82 (0.73–0.91)	0.0901	0.89 (0.79–1.02)
rs11175885	12	64,400,897	*C12orf66/TMEM5*	3.66E-04	0.83 (0.75–0.92)	0.1247	0.90 (0.80–1.03)
rs2358944	12	64,403,825	*MSRB3/HMGA2*	1.15E-04	0.82 (0.74–0.90)	0.1100	0.90 (0.79–1.02)
rs2904532	12	65,582,311	*GRIP1/CAND1*	4.13E-04	0.82 (0.73–0.91)	0.0902	0.89 (0.77–1.02)
rs11176482	12	65,582,814	*GRIP1/CAND1*	N/A	N/A	N/A	N/A
rs12302041	12	76,654,356	*NAV3*	1.87E-04	1.40 (1.17–1.67)	0.0895	1.21 (0.97–1.51)
rs1978243	14	86,596,160	No gene	6.25E-04	0.84 (0.76–0.93)	0.2588	0.93 (0.83–1.05)
rs6494387	15	61,132,579	*TPM1*	0.0017	1.16 (1.06–1.27)	0.4478	1.04 (0.93–1.17)
rs1034589	22	29,909,233	*RNF185*	2.49E-08	0.56 (0.46–0.69)	2.17E-06	0.52 (0.40–0.68)
rs2106294	22	29,975,759	*LIMK2*	3.28E-08	0.56 (0.46–0.69)	7.15E-06	0.54 (0.42–0.71)
rs4820043	22	29,977,094	*LIMK2*	3.04E-08	0.56 (0.46–0.69)	5.77E-06	0.54 (0.41–0.70)
rs5749286	22	30,230,359	*SFI1*	6.51E-11	0.51 (0.41–0.62)	1.97E-10	0.38 (0.28–0.51)
rs16996381	22	34,841,225	*APOL3*	2.96E-07	0.80 (0.73–0.87)	7.20E-07	0.76 (0.69–0.85)
rs735853	22	35,009,161	*MYH9*	4.77E-10	0.63 (0.54–0.73)	7.76E-10	0.55 (0.45–0.66)

Abbreviations: Chr., chromosome; ESRD, end-stage renal disease; N/A, not applicable; OR, odds ratio; T2DM, type 2 diabetes mellitus; 95% CI, 95% confidence interval.
SNP rs11176482 failed to type in non-T2DM-ESRD subjects.
Comparison of all-cause ESRD (1674 T2DM-ESRD cases and 1216 non-T2DM-ESRD cases) with controls (combined from GWAS+replication), and non-T2DM-ESRD cases with controls.

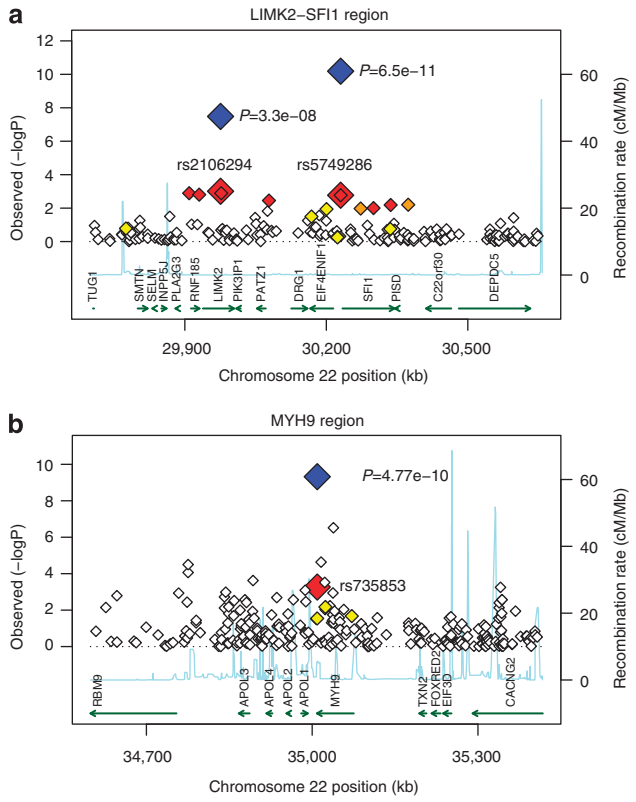

Figure 2 | All-cause ESRD candidate regions. (**a**) LIMK2-SFI1 gene region. (**b**) MYH9 gene region. $-\log_{10} P_{ADD}$ values from the GWAS are plotted versus position (genome build 36). The large red diamond indicates P_{ADD} values from the GWAS of the marker(s) displayed. The large blue diamond and corresponding *P*-value indicates P_{ADD} values from the combined analysis of the marker(s) displayed. LD based on control samples is color coded: red (r^2 to top SNP 0.8–1.0), orange (0.5–0.8), yellow (0.2–0.5), and white (<0.2). Gene annotations were obtained from UCSC Genome Browser (RefSeq Genes, b36). Arrows represent direction of transcription. ESRD, end-stage renal disease; GWAS, genome-wide association study; LD, linkage disequilibrium; MYH9, nonmuscle myosin heavy chain 9.

strongly associated in the test of all-cause ESRD (Table 5), and only four of these markers (rs1034589, rs5749286, rs16996381, and rs735853) showed stronger association with nondiabetic ESRD (Table 5) than with DN (Table 3, combined analysis).

The four hits in the *LIMK2-SFI1* region are all highly correlated (Supplementary Table S4 online). SNPs rs1034589 in *RNF185*, rs2106294 in *LIMK2*, and rs4820043 also in *LIMK2* are all perfectly correlated with one another ($r^2 = 1.0$, Supplementary Table S4 online). In addition, these SNPs are all highly correlated with rs5749286 in *SFI1* ($r^2 = 0.85$, Supplementary Table S4 online). However, the *LIMK2-SFI1* region and the *MYH9* region are poorly correlated with r^2 values ranging from 0.0324 to 0.0335 (Supplementary Table S4 online).

Tests for association of previously identified nephropathy genes

In previous studies,[7,8,10–12,18–23] multiple loci have been implicated in different ethnicities with either renal failure or renal function in both diabetic and nondiabetic patients. We have assessed association with many of these genes (such as *PVT1*, *FRMD3*, *CARS*, *ACACB*, *NEDD4L*, *SERPINB7*, *CNDP1*, *CNDP2*, *ELMO1*, *SHROOM3*, *UMOD*, *GATM-SPATA5L1*, *GCK2*, *ALMS1*, *DAB2*, *SLC34A1*, *VEGFA*, *SLC22A2*, *PRKAG2*, *STC1*, *ATXN2*, *DACH1*, *SLC7A9)* and T2DM-ESRD in our GWAS. For each locus, the entire genic region, along with 10 kb upstream and downstream of the gene, was inspected for association in the GWAS data to see whether other variants in the same gene were associated with T2DM-ESRD in AAs. All of these results are presented in Supplementary Table S5 online. Although there are some results that are nominally significant, none of the results meet a Bonferroni correction ($P = 0.05/1909$ total SNPs $= 2.6 \times 10^{-5}$) with the lowest *P*-value for this analysis being 5.8×10^{-4} in the *NEDD4L* gene.

DISCUSSION

We performed a high-density GWAS to investigate genetic susceptibility to T2DM-ESRD and all-cause ESRD in AAs. Previous GWASs for diabetes-associated nephropathy in Japanese,[7] Pima Indians,[8] and CAs [9] have identified several potential DN loci: *ELMO1*,[7] *PVT1*,[8] *FRMD3*, and *CARS*.[9] We have previously observed association with the *ELMO1* and *MYH9* genes and T2DM-ESRD in our AA population.[17,20] These studies in other ethnicities have used a study design of comparing cases with T2DM-ESRD and controls with T2DM and no nephropathy. Although this study design might be appropriate for other ethnicities, it has serious limitations for AAs. AAs with T2DM and preserved renal function for ⩾10 years ('hypernormal' controls) are perhaps the ideal control group, but are rare.[24] We estimate that <10% of AA T2DM affected patients of 10 years duration have preserved renal function. We have chosen a study design that is flexible and support studies of T2DM and body mass index in parallel with T2DM-ESRD. Importantly, we have used both replication and several contrast groups to differentiate between T2DM-ESRD and T2DM. This system works quite well: the strongest T2DM-associated gene in AAs is *TCF7L2*, which is easily filtered out of our hits for T2DM-ESRD, but prominently associated with T2DM.[25,26]

We identified five novel gene regions with evidence of association with T2DM-ESRD in AAs ($P < 1.0 \times 10^{-5}$): *SASH1*, *RPS12*, *AUH*, *MSRB3-HMGA2*, and *LIMK2-SFI1*. One of these regions, *LIMK2-SFI1*, was also significantly associated ($P < 5.0 \times 10^{-8}$) with all-cause ESRD. *MYH9*, which has been previously associated with nondiabetic nephropathy,[14–16] and DN,[17] also showed significant association ($P < 5.0 \times 10^{-8}$) with all-cause ESRD.

SASH1, located at 6q24.3, encodes the sterile α-motif (SAM)- and SH3-domain containing 1 protein and is ubiquitously expressed. *SASH1* was originally identified as a candidate tumor-suppressor gene, being downregulated in the majority (74%) of breast cancer tumors,[27] and has been shown to be downregulated in colon cancer.[28] SASH1 also functions as a signaling adapter, as a downstream target in

the insulin-like growth factor-1/phosphatidylinositol 3-kinase signaling pathway.[29] SNP rs6930576, located within this gene, showed the strongest association with T2DM-ESRD in the combined analysis, suggesting that this gene may also have a role in the progression of DN.

Two other candidate regions are also worth mentioning. The ribosomal protein S12 gene (*RPS12*) is located on chromosome 6q23.2. The protein encoded by this gene is located in the cytoplasm and is a component of the 40S ribosomal subunit.[30] *RPS12* is also a potential target gene of microRNA-377, which has been consistently upregulated in *in vitro* DN models and in *in vivo* DN mouse models.[31] If *RPS12* is also upregulated in the diabetic milieu, it may contribute to the progression of DN.

LIMK2, LIM kinase 2, is located on chromosome 22 at 22q12.2.[32] There are three isoforms of *LIMK2* produced by alternative splicing: variants 1, 2a, and 2b.[33] *LIMK2* is one of two members of the LIM kinase family, and contains two LIM domains at the N terminus and a protein kinase domain at the C terminus.[32] LIMK2 is activated by Rho and Cdc42 by ROCK and PAK, respectively; in turn, LIMK2 phosphorylates cofilin, eliminating cofilin's ability to bind and depolymerize actin. This leads to an accumulation of actin filaments and aggregates. It is through this action that LIMK2 has a role in the Rho- and Cdc42-induced reorganization of the actin cytoskeleton.[34–36] LIMK2 is widely expressed in all tissues,[37] showing moderate expression in the kidney.[32,37] Homozygous Limk2 gene-deficient mice (Limk2$^{-/-}$), created using a Cre-mediated excision to disrupt all three LIMK2 isoforms, exhibited normal growth to adulthood when compared with wild-type and heterozygous mice.[38] However, Limk2$^{-/-}$ mice had moderate kidney abnormalities with dilated collecting tubules and slight decreases in the number of glomeruli, without changes in nephron function.[38] If variants in *LIMK2* lead to kidney abnormalities in humans, this may be detrimental in a diabetic environment, and could possibly lead to progression of DN. This would also support *LIMK2* as a potential all-cause ESRD candidate locus.

The other gene in the *LIMK2* region that showed evidence of association with T2DM-ESRD and all-cause ESRD was *SFI1*, Sfi1 homolog, spindle assembly-associated (yeast). In yeast, SFI1 has been shown to be an essential and conserved component of centrosomes—known as spindle pole bodies in yeast.[39] In addition, SFI1 was shown to be required for spindle pole body duplication and may be required for separation of the duplicated spindle pole bodies.[40]

MYH9, which showed evidence of association with T2DM-ESRD and all-cause ESRD in our study, has been previously associated with non-T2DM nephropathy[14–16] and with diabetic forms of nephropathy.[17] It should be noted that the results reported in this study are in reference to the minor (less frequent) allele, and are associated with protection. This remains consistent with previous associations in which the reported risk allele corresponded to the major allele.[14–17] In Yoruba (YRI) samples in HapMap, the protective allele for rs735853 is observed in only 3% of chromosomes, whereas it is seen in >57% of CEU chromosomes. For our AA sample, the allele frequencies are 0.082 for cases and 0.123 for controls. This result may reflect a selection for the admixed European allele, or more broadly, for the European locus, which provides protection in AAs. Whether this reflects that some cases in our study could be non-diabetic ESRD (Freedman *et al.*[17]: recently re-estimated to be 13%) or reflects a role for this allele in T2DM-ESRD remains to be determined. Parenthetically, the *LIMK2-SFI1* gene region is located ~5 Mb away from the *MYH9* gene region. We evaluated linkage disequilibrium across chromosome 22 (Supplementary Table S4 online). The SNPs that showed evidence of association in the *LIMK2-SFI1* gene region were poorly correlated ($r^2 \leqslant 0.03$) with the SNP that showed evidence of association in *MYH9*, suggesting that these are independent signals.

However, *MYH9* and *LIMK2* are both involved in the reorganization and maintenance of the actin cytoskeleton. *MYH9* encodes the nonmuscle myosin heavy chain IIA, which is highly conserved and similar to other nonmuscle myosins.[41,42] Myosin IIA is expressed in podocytes and mesangial cells.[41] Within the podocyte, myosin IIA localizes to the foot process in which it is involved in the movement of actin filaments to maintain cell structure.[43] Many studies have shown that mutations in *MYH9* are associated with kidney disease.[14–17] As LIMK2 is in the same pathway, it remains possible that mutations in *LIMK2* could produce kidney disease; thus, *LIMK2* is a strong candidate in T2DM-ESRD and all-cause ESRD.

We performed two further assessments of loci in this study. First, we assessed several of the AA loci (13/25) to test whether they contributed to T2DM-ESRD in a sample of CA cases and controls. We observed no evidence for association in CAs (data not shown), but a more comprehensive, better-powered experiment might be more definitive. Conversely, we did not observe evidence of association with any of the DN and chronic kidney disease candidate genes and T2DM-ESRD in AAs in the current study. The GWAS herein did not interrogate some of the specific disease associated SNPs such as *ACACB* SNP rs2268388 with which we previously reported evidence of association.[21] The large number of SNPs evaluated in this study (>1900) results in a more stringent threshold for significance. As these variants were mainly identified in non-African derived racial/ethnic groups, they may have a limited role in the development of T2DM-ESRD in AAs.

This study has similar limitations as other GWAS. Although we have identified 25 variants that may influence T2DM-ESRD in 19 candidate regions, none of these regions have reached genome-wide significance ($P < 5.0 \times 10^{-8}$) with the T2DM-ESRD phenotype. However, even though this study has a limited sample size for the GWAS discovery phase, power calculations (Supplementary Table S6 online) show that we have moderate statistical power to detect ORs of 1.3 (minor allele frequency = 0.35), consistent with previously published effect sizes for DN (*FRMD3* and *CARS*,

OR = 1.45 and OR = 1.36, respectively[9]). In addition, other phases of the study: replication, trait discrimination, and all-cause ESRD analysis have excellent power for detecting association, given the goals of each step (for example, nominal association in replication; P-value $\leqslant 1 \times 10^{-6}$ for the all-cause ESRD analysis). There are almost certainly additional loci of moderate effect that have not been captured in this study, and loci of weaker effect will require a significantly larger study. There are few other existing collections of appropriate samples in AAs, and we are one of the few centers with active recruiting efforts in this population. This will make the search for additional replication populations more difficult. In addition, it is difficult to ensure that all subjects clinically labeled as having T2DM-ESRD lack other renal lesions. This difficulty plagues all studies of DN and we chose inclusion criteria similar to other large studies.[44]

In conclusion, we performed a GWAS for T2DM-ESRD in AAs. We then carried out a replication phase and a trait discrimination phase to determine whether associations were with T2DM, T2DM-ESRD, and/or all-cause ESRD. Through these studies, we have discovered 25 potential variants in 19 genes associated with T2DM-ESRD, including *SASH1*, *RPS12*, *AUH*, *MSRB3-HMGA2*, and *LIMK2-SFI1*. These results require further replication to confirm their role in T2DM-ESRD susceptibility and to clarify their role in all-cause ESRD in AAs.

MATERIALS AND METHODS

Subjects

GWAS samples and clinical characteristics. Recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University, and informed consent was obtained from all study participants. Patients with T2DM were recruited from dialysis facilities. Individuals with a history of ketoacidosis or those who developed diabetes before the age of 25 years and received continuous insulin therapy since diagnosis were assumed to have type 1 diabetes and were excluded. T2DM was diagnosed in AAs who reported developing diabetes after the age of 25 years and who did not receive only insulin therapy since diagnosis. Cases had T2DM diagnosed at least 5 years before initiating renal replacement therapy, background or greater diabetic retinopathy, and/or $\geqslant$100 mg/dl proteinuria on urinalysis in the absence of other causes of nephropathy (T2DM-ESRD subjects). Unrelated AA controls without a current diagnosis of diabetes or renal disease were recruited from the community and internal medicine clinics (control subjects). All cases and nondiabetic, nonnephropathy controls were born in North Carolina, South Carolina, Georgia, Tennessee, or Virginia. DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN).

Replication study samples and clinical characteristics. Informed consent was obtained from all study participants, and recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University. AA T2DM-ESRD cases and nondiabetic, nonnephropathy controls were recruited using the same criteria as the case and control subjects that were used in the GWAS. DNA extraction was performed using the PureGene system.

T2DM, nonnephropathy and non-T2DM-ESRD study samples, and clinical characteristics. Sample collection procedures and recruitment were approved by the Wake Forest University Institutional Review Board. Informed consent was obtained from all study participants. Subjects with T2DM without evidence of nephropathy were recruited from medical clinics, churches, health fairs, and community resources (T2DM subjects). Patients who had ESRD attributed to hypertension or primary glomerular disease were recruited (non-T2DM-ESRD subjects). Putative hypertension-associated ESRD was diagnosed in patients with high blood pressure preceding the initiation of renal replacement therapy and low-level proteinuria ($\leqslant$30 mg/dl on urine dipstick, $<$0.5 g protein/24 h on timed urine collection, or urine protein: creatinine ratio $<$0.5 g/g) or in the absence of measures of proteinuria. Chronic glomerular disease-associated ESRD was diagnosed in nondiabetic subjects with renal biopsy evidence of primary glomerular disease (for example, focal segmental glomerulosclerosis), proteinuria $\geqslant$0.5 g/24 h, or $\geqslant$100 mg/dl on urinalysis. Individuals with DN (as described above), cystic renal diseases, hereditary nephritis, or urologic causes of ESRD were excluded. Individuals were unrelated and self-described AA. All subjects were born in North Carolina, South Carolina, Georgia, Virginia, or Tennessee. The PureGene system was used for DNA extraction.

Sample preparation, genotyping, and quality control

Genome-wide association study. Genotyping was performed at the Center for Inherited Disease Research (CIDR) using 1 μg of genomic DNA (diluted in 1 × TE buffer and at 50 ng/μl) on the Affymetrix Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA). DNA from cases and controls were equally interleaved on 96-well master plates to ensure technical uniformity during sample processing. To confirm sample identity, an SNP barcode (96 SNPs) was generated before genotyping on the Affymetrix array and was confirmed on downstream-released genotyping data. Genotypes were called using Birdseed version 2; Affymetrix Power Tools (APT) 1.10.0 (Affymetrix) by grouping samples by DNA plate to determine the genotype cluster boundaries. All autosomal SNPs (n = 868,157) were included in the analysis but were classified on data quality. The primary inference was drawn from SNPs that had $<$5% missing data, Hardy–Weinberg P-values in cases $>$0.0001 and in controls $>$0.01, no significant difference in missing data rate between cases and controls and were polymorphic (n = 832,357). The average sample call rate was 99.16% for all autosomal SNPs. A total of 46 blind duplicates were included in genotyping and had a concordance rate of 99.59%. In addition, individuals whose gender call from X-chromosome genotype data was discordant with the gender obtained from patient interviews were excluded from the analysis (n = 1). Cryptic relatedness was estimated by pairwise identity-by-descent analysis implemented in the PLINK analysis software package (http://pngu.mgh.harvard.edu/purcell/plink/).[45] Two duplicate samples were identified, and one sample in each duplicate pair was removed. In addition, 104 individuals from pairs of samples were identified as cryptic first-degree relatives; these individuals were kept for the analysis. We also assessed heterozygosity by estimating the inbreeding coefficient, F, using PLINK. One subject had an F-value $>$ 4 s.d. from the mean, suggesting population substructure. This subject was removed. Our final data set consisted of 1994 individuals (965 T2DM-ESRD cases and 1029 controls) in which we performed association analysis.

Replication. The replication population was recruited under identical ascertainment criteria to that of the GWAS. A total of 724 SNPs were genotyped using the iPLEX Sequenom MassARRAY platform

(Sequenom, San Diego, CA). Genotyping efficiency was >90% and 45 blind duplicates were included to ensure genotyping accuracy.

Trait discrimination. A total of 67 SNPs were genotyped using the iPLEX Sequenom MassARRAY platform. Genotyping efficiency was >90%. A total of 50 blind duplicates were included to ensure genotyping accuracy.

Analysis

Genome-wide association study. To address the effect of admixture in this AA data set, we performed a principal components analysis, which used all high-quality data from the GWAS and excluded regions of high linkage disequilibrium and inversions. This approach is an iterative process whereby high-quality autosomal SNPs ($n = 832{,}357$) are used to calculate the top 50 principal components. Once calculated, the principal components are examined to determine whether they are tied to a region of the genome. If so, those SNPs are excluded and the analysis is repeated. The first principal component (PC1) explained the largest proportion of variation at 22%. A direct comparison of principal components analysis with frequenist estimation of individual ancestry proportion analysis of 70 ancestry informative markers resulted in a high correlation between PC1 and the ancestry informative markers ($r^2 = 0.87$). Other principal components were associated with regions of the genome, representing another unclassified source of variance, and did not further reduce the inflation factor. The mean (s.d.) African ancestry proportions in 965 T2DM-ESRD cases and 1029 controls were 0.80 ± 0.11 and 0.78 ± 0.11, respectively, as estimated by Frequentist Estimation of individual ancestry proportion analysis.[46] To test for association with T2DM-ESRD, genotypic tests of association were performed on each SNP individually using SNPGWA version 4.0 (http://www.phs.wfubmc.edu),[47] an analytic package that includes the capability to perform association calculations adjusting for covariates. The genotypic association reported in this study is for analyses incorporating adjustment for PC1. The primary inference is based on the additive genetic model. The inflation factor was calculated from the observed mean χ^2 statistic across high-quality autosomal SNPs ($n = 832{,}357$). Under the null hypothesis, if there is no inflation, the expected mean value is 1. In our population, the inflation factor calculated in the GWAS sample with 104 cryptic-related individuals ($n = 1994$), adjusting for PC1 was 1.06. After 54 subjects were removed to eliminate the cryptic relations ($n = 1940$), the inflation factor in the GWAS, adjusting for PC1 was 1.04. With this minimal improvement in inflation, all subjects ($n = 1994$) were retained for analysis as the difference between the inflation factors is only 0.02.

Replication in T2DM-ESRD cases and nondiabetic, nonnephropathy controls. To account for admixture in the replication cohort, ancestral allele frequencies were estimated from the results of the 70 ancestry informative markers genotyped in 44 Yoruba Nigerians and 39 European Americans. Individual ancestral proportions were generated for each subject using frequenist estimation of individual ancestry proportion, an EM algorithm, under a two-population model. The mean (s.d.) African ancestry proportions in 709 T2DM-ESRD cases and 690 controls were 0.80 ± 0.12 and 0.76 ± 0.13, respectively. Association analysis was performed as described for the GWAS.

Combined analysis of T2DM-ESRD cases and nondiabetic, nonnephropathy controls. The GWAS and replication cohorts were merged as one cohort for a combined analysis. Admixture in the combined cohort was accounted for by using ancestral allele frequencies as described for the replication analysis. The association analysis was performed as described for the GWAS and replication study.

Discrimination between T2DM-ESRD, T2DM, and all-cause ESRD. To determine the association with T2DM-ESRD, T2DM, and/or all-cause ESRD, additional association analyses were performed as described for the GWAS, replication, and combined analysis incorporating the 1246 T2DM, nonnephropathy subjects and the 1216 non-T2DM-ESRD subjects. Admixture was accounted for through ancestral allele frequencies as described in the replication and combined analysis. The mean (s.d.) African ancestry proportions in 1246 T2DM, nonnephropathy cases and 1216 non-T2DM-ESRD cases were 0.76 ± 0.13 and 0.79 ± 0.12, respectively.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

We thank the patients, their relatives, and the staff of the Southeastern Kidney Council/ESRD Network 6 for their participation. This work was supported by NIH grants R01 DK066358 (DWB), R01 DK053591 (DWB), R01 HL56266 (BIF), R01 DK070941 (BIF), and in part by the General Clinical Research Center of the Wake Forest University School of Medicine grant M01 RR07122. This work was presented in abstract form at the 2009 Annual Meeting of the American Society of Human Genetics, Honolulu, HI, 20–24 October; at the 2009 Annual Meeting of the American Society of Nephrology, San Diego, CA, 29 October through 1 November; and at the 2010 Annual Meeting of the American Diabetes Association, Orlando, FL, 25–29 June.

SUPPLEMENTARY MATERIAL

Table S1. GWAS, replication, and combined (GWAS + replication) *P*-values for 724 SNPs selected for replication across the genome.
Table S2. GWAS, replication, and combined *P*-values for SNPs in and near *MYH9* on chromosome 22.
Table S3. Trait discrimination analyses and tests for all-cause ESRD for the 67 SNPs that passed replication across the genome.
Table S4. Measures of linkage disequilibrium across the diabetic nephropathy loci on chromosome 22.
Table S5. Evaluation of known diabetic nephropathy and chronic kidney disease candidates in the genome-wide association study.
Table S6. Power estimates for each stage.
Figure S1. Quantile–quantile plot of the genome-wide association study results.
Figure S2. Diabetic nephropathy candidate regions.
Supplementary material is linked to the online version of the paper at http://www.nature.com/ki

REFERENCES

1. USRDS: US Renal Data System 2009 Annual Data Report. *Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. In: National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases: Bethesda, MD, 2009.
2. Bowden DW. Genetics of kidney disease. *Kidney Int Suppl* 2003; **83**: S8–12.
3. Quinn M, Angelico MC, Warram JH *et al.* Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* 1996; **39**: 940–945.
4. Seaquist ER, Goetz FC, Rich S *et al.* Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med* 1989; **320**: 1161–1165.
5. Spray BJ, Atassi NG, Tuttle AB *et al.* Familial risk, age at onset, and cause of end-stage renal disease in White Americans. *J Am Soc Nephrol* 1995; **5**: 1806–1810.
6. Freedman BI, Tuttle AB, Spray BJ. Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. *Am J Kidney Dis* 1995; **25**: 710–713.

7. Shimazaki A, Kawamura Y, Kanazawa A *et al.* Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes* 2005; **54**: 1171–1178.
8. Hanson RL, Craig DW, Millis MP *et al.* Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. *Diabetes* 2007; **56**: 975–983.
9. Pezzolesi MG, Poznik GD, Mychaleckyj JC *et al.* Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 2009; **58**: 1403–1410.
10. Kottgen A, Glazer NL, Dehghan A *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* 2009; **41**: 712–717.
11. Kottgen A, Pattaro C, Boger CA *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* 2010; **42**: 376–384.
12. Chambers JC, Zhang W, Lord GM *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* 2010; **42**: 373–375.
13. CDC: National Diabetes Fact Sheet: general information and national estimates on diabetes in the United States, 2007 In: *US Department of Health and Human Services*, edited by CDC: Atlanta, GA, 2008.
14. Kopp JB, Smith MW, Nelson GW *et al.* MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* 2008; **40**: 1175–1184.
15. Kao WH, Klag MJ, Meoni LA *et al.* MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet* 2008; **40**: 1185–1192.
16. Freedman BI, Hicks PJ, Bostrom MA *et al.* Polymorphisms in the non-muscle myosin heavy chain 9 gene (MYH9) are strongly associated with end-stage renal disease historically attributed to hypertension in African Americans. *Kidney Int* 2009; **75**: 736–745.
17. Freedman BI, Hicks PJ, Bostrom MA *et al.* Non-muscle myosin heavy chain 9 gene MYH9 associations in African Americans with clinically diagnosed type 2 diabetes mellitus-associated ESRD. *Nephrol Dial Transplant* 2009; **24**: 3366–3371.
18. Freedman BI, Hicks PJ, Sale MM *et al.* A leucine repeat in the carnosinase gene CNDP1 is associated with diabetic end-stage renal disease in European Americans. *Nephrol Dial Transplant* 2007; **22**: 1131–1135.
19. Janssen B, Hohenadel D, Brinkkoetter P *et al.* Carnosine as a protective factor in diabetic nephropathy: association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes* 2005; **54**: 2320–2327.
20. Leak TS, Perlegas PS, Smith SG *et al.* Variants in intron 13 of the ELMO1 gene are associated with diabetic nephropathy in African Americans. *Ann Hum Genet* 2009; **73**: 152–159.
21. Maeda S, Kobayashi M, Araki S *et al.* A single nucleotide polymorphism within the acetyl-coenzyme a carboxylase beta gene is associated with proteinuria in patients with type 2 diabetes. *Plos Genet* 2010; **6**: e1000842.
22. McDonough CW, Bostrom MA, Lu L *et al.* Genetic analysis of diabetic nephropathy on chromosome 18 in African Americans: linkage analysis and dense SNP mapping. *Hum Genet* 2009; **126**: 805–817.
23. McDonough CW, Hicks PJ, Lu L *et al.* The influence of carnosinase gene polymorphisms on diabetic nephropathy risk in African-Americans. *Hum Genet* 2009; **126**: 265–275.
24. Gitter J, Langefeld CD, Rich SS *et al.* Prevalence of nephropathy in black patients with type 2 diabetes mellitus. *Am J Nephrol* 2002; **22**: 35–41.
25. Lewis JP, Palmer ND, Hicks PJ *et al.* Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes* 2008; **57**: 2220–2225.
26. Sale MM, Smith SG, Mychaleckyj JC *et al.* Variants of the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in an African-American population enriched for nephropathy. *Diabetes* 2007; **56**: 2638–2642.
27. Zeller C, Hinzmann B, Seitz S *et al.* SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is downregulated in breast cancer. *Oncogene* 2003; **22**: 2972–2983.
28. Rimkus C, Martini M, Friederichs J *et al.* Prognostic significance of downregulated expression of the candidate tumour suppressor gene SASH1 in colon cancer. *Br J Cancer* 2006; **95**: 1419–1423.
29. Dubois F, Vandermoere F, Gernez A *et al.* Differential 14-3-3 affinity capture reveals new downstream targets of phosphatidylinositol 3-kinase signaling. *Mol Cell Proteomics* 2009; **8**: 2487–2499.
30. Yu Y, Ji H, Doudna JA *et al.* Mass spectrometric analysis of the human 40S ribosomal subunit: native and HCV IRES-bound complexes. *Protein Sci* 2005; **14**: 1438–1446.
31. Wang Q, Wang Y, Minto AW *et al.* MicroRNA-377 is up-regulated and can lead to increased fibronectin production in diabetic nephropathy. *FASEB J* 2008; **22**: 4126–4135.
32. Okano I, Hiraoka J, Otera H *et al.* Identification and characterization of a novel family of serine/threonine kinases containing two N-terminal LIM motifs. *J Biol Chem* 1995; **270**: 31321–31330.
33. Scott RW, Olson MF. LIM kinases: function, regulation and association with human disease. *J Mol Med* 2007; **85**: 555–568.
34. Sumi T, Matsumoto K, Takai Y *et al.* Cofilin phosphorylation and actin cytoskeletal dynamics regulated by rho- and Cdc42-activated LIM-kinase 2. *J Cell Biol* 1999; **147**: 1519–1532.
35. Edwards DC, Sanders LC, Bokoch GM *et al.* Activation of LIM-kinase by Pak1 couples Rac/Cdc42 GTPase signalling to actin cytoskeletal dynamics. *Nat Cell Biol* 1999; **1**: 253–259.
36. Ohashi K, Nagata K, Maekawa M *et al.* Rho-associated kinase ROCK activates LIM-kinase 1 by phosphorylation at threonine 508 within the activation loop. *J Biol Chem* 2000; **275**: 3577–3582.
37. Acevedo K, Moussi N, Li R *et al.* LIM kinase 2 is widely expressed in all tissues. *J Histochem Cytochem* 2006; **54**: 487–501.
38. Takahashi H, Koshimizu U, Miyazaki J *et al.* Impaired spermatogenic ability of testicular germ cells in mice deficient in the LIM-kinase 2 gene. *Dev Biol* 2002; **241**: 259–272.
39. Kilmartin JV. Sfi1p has conserved centrin-binding sites and an essential function in budding yeast spindle pole body duplication. *J Cell Biol* 2003; **162**: 1211–1221.
40. Anderson VE, Prudden J, Prochnik S *et al.* Novel sfi1 alleles uncover additional functions for Sfi1p in bipolar spindle assembly and function. *Mol Biol Cell* 2007; **18**: 2047–2056.
41. Arrondel C, Vodovar N, Knebelmann B *et al.* Expression of the nonmuscle myosin heavy chain IIA in the human kidney and screening for MYH9 mutations in Epstein and Fechtner syndromes. *J Am Soc Nephrol* 2002; **13**: 65–74.
42. D'Apolito M, Guarnieri V, Boncristiano M *et al.* Cloning of the murine non-muscle myosin heavy chain IIA gene ortholog of human MYH9 responsible for May-Hegglin, Sebastian, Fechtner, and Epstein syndromes. *Gene* 2002; **286**: 215–222.
43. Faul C, Asanuma K, Yanagida-Asanuma E *et al.* Actin up: regulation of podocyte structure and function by components of the actin cytoskeleton. *Trends Cell Biol* 2007; **17**: 428–437.
44. Knowler WC, Coresh J, Elston RC *et al.* The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. *J Diabetes Complicat* 2005; **19**: 1–9.
45. Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *AM J Hum Gen* 2007; **81**: 559–575.
46. Tang H, Peng J, Wang P *et al.* Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 2005; **28**: 289–301.
47. Harley JB, Alarcon-Riquelme ME, Criswell LA *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008; **40**: 204–210.